Integration of audio-visual information for spatial decisions in children and adults

Marko Nardini[1,2]*, Jennifer Bales*[2] *and Denis Mareschal*[3]

[1] Dept of Psychology, University of Durham, South Road, Durham DH1 3LE, UK.

[2] Dept of Visual Neuroscience, UCL Institute of Ophthalmology, London, EC1V 9EL, UK.

[2] Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck University of London, London WC1E 7HX, UK

In press, *Developmental Science* (2015)

Corresponding author: marko.nardini@durham.ac.uk, +44 (0) 191 334 3255

Research highlights

- When judging the locations of targets, children aged 4 to 12 years and adults had faster and less variable response times given auditory and visual information together compared with either alone.

- Comparison of response time distributions with model predictions indicates that children at all ages were integrating (pooling) sensory information to make decisions. However, both overall speed and the efficiency of sensory integration improved with age.

- This evidence for pooling comes not only from the standard approach of comparison with Miller's seminal "race model", but also from comparison with a major recent extension of this model and with a comparable "pooling" (coactivation) model.

- Our results, which come from a wide age range tested with the same task, may reconcile previous results using different tasks from which the developmental time course of audio-visual integration was unclear.

In adults, decisions based on multisensory information can be faster and/or more accurate than those relying on a single sense. However, this finding varies significantly across development. Here we studied speeded responding to audio-visual targets, a key multisensory function whose development remains unclear. We found that when judging the locations of targets, children aged 4 to 12 years and adults had faster and less variable response times given auditory and visual information together compared with either alone. Comparison of response time distributions with model predictions indicated that children at all ages were integrating (pooling) sensory information to make decisions but that both the overall speed and the efficiency of sensory integration improved with age. The evidence for pooling comes from comparison with the predictions of Miller's seminal "race model", as well as with a major recent extension of this model and a comparable "pooling" (coactivation) model. The findings and analyses can reconcile results from previous audio-visual studies, in which infants showed speed gains exceeding race model predictions in a spatial orienting task (Neil et al, *Dev Sci* 2006) but children below 7 years did not in speeded reaction time tasks (*e.g.* Barutchu et al, *Dev Sci* 2009). Our results provide new evidence for early and sustained abilities to integrate visual and auditory signals for spatial localisation from a young age.

Keywords: multisensory, visual, auditory, development

Events in the world are often signalled to more than one sense; for example, we can both see and hear an approaching car. The availability of parallel information sources provides several potential advantages to sensory systems. In tasks requiring estimates of the size or position of noisy stimuli, weighted averaging of multiple estimates can reduce the uncertainty of the final estimate (Clark & Yuille, 1990; Ernst, 2005). In tasks requiring speeded responses to supra-threshold stimuli, parallel processing of multiple signals predicts some speed gains (Raab, 1962), and pooling of signals even greater speed gains (Miller, 1982; Schroger & Widmann, 1998; Townsend & Wenger, 2004). How these kinds of sensory optimizations emerge over the course of human development is currently unclear. Although multisensory links develop early in childhood (*e.g.* Lewkowicz & Turkewitz, 1981; Bahrick & Lickliter, 2000), abilities to improve perceptual precision by integrating multiple sensory cues have not been observed until after the age of 8 years (Gori, Del Viva, Sandini, & Burr, 2008; Nardini, Jones, Bedford, & Braddick, 2008; Nardini, Bedford, & Mareschal, 2010; Gori, Sandini, & Burr, 2012; Nardini, Begus, & Mareschal, 2013; Petrini, Remark, Smith, & Nardini, 2014). In tasks measuring response speed, development with age of advantages for multi- *vs* uni-sensory stimuli has been inconsistent across different tasks (Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006; Barutchu, Crewther, & Crewther, 2009).

Here we studied the development of speeded localisation of audio-visual targets from 4 years. The development of abilities to speed up response times given parallel auditory and visual information is currently unclear. A central question for behavioural studies is whether multisensory response time advantages are within the range that can be explained by the seminal "race models" which have been used to predict gains that can be simply due to parallel processing of two signals (Raab, 1962; Miller, 1982). Because of trial-to-trial variability in response times based on any single channel, an observer who responds based on the fastest single channel on each trial (i.e. *via* a

"race" between signals) can make predictable speed gains. When speed gains exceed those predicted by a race, they have been interpreted as arising instead from the pooling of signals by a multisensory decision mechanism (Miller, 1982; Schroger & Widmann, 1998). At 8-10 months, infants orienting spontaneously to audio-visual targets have shown speed gains exceeding those of a race model (Neil *et al.*, 2006). In contrast, when children had to press a button in response to a central target, the ability to out-perform a race model first emerged much later, at 7 years (Barutchu *et al.*, 2009) or 10-12 years (Brandwein *et al.*, 2011), and was not present at all older ages (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010). One potentially important difference between these studies is that infants, but not older children, were moving their eyes and head to localise stimuli whose spatial locations changed from trial to trial.

In addition, a recent re-formulation of the classic race model has shown that some speed gains relative to the classic model can in fact still be explained by parallel processing, if the need to process two signals adds variability to decision times (Otto & Mamassian, 2012). Importantly, this calls into question previous evidence for multisensory integration based on the classic model (*e.g.* Schroger & Widmann, 1998; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994), including all the developmental studies to date (Neil *et al.*, 2006; Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011; Barutchu *et al.*, 2011). The course of development of these mechanisms is therefore extremely unclear.

An important neural substrate for auditory-visual integration is the superior colliculus (SC), associated with spatial localisation for oculomotor and orienting responses (Stein & Meredith, 1993). Multisensory receptive fields of SC neurons are known from animal models to develop postnatally in an experience-dependent manner (Wallace, Wilkinson, & Stein, 1996; Wallace & Stein, 1997; Stein, Stanford, & Rowland, 2009). These neurons may support pooling of auditory

and visual information, as they respond to either signal, and can show "super-additive" properties in responding more strongly to both signals together than to either alone. Additional, cortical mechanisms for auditory-visual integration have also been described in human adults (Molholm *et al.*, 2002; Martuzzi *et al.*, 2007; Noppeney, Ostwald, & Werner, 2010). A developmental behavioural and ERP study (Brandwein *et al.*, 2011) found speed gains exceeding race model predictions only in groups aged 10 years and above (but not 7-9 years), and an overall correlation between behavioural speed gains and differences in summed single-cue and combined-cue evoked potentials at a fronto-central scalp region. This illustrates the contribution of at least one cortical mechanism to the development of audio-visual integration. There are likely to be multiple auditory-visual integration mechanisms, sub-cortical and cortical, whose relative contributions to multisensory processing may be task-dependent, and could develop at different rates in childhood. We reasoned that speed gains should be particularly evident for spatial localisation of targets, a task which could tap into the spatial coding provided by the superior colliculus, once this system for integration of spatial auditory and visual signals has developed. To study speeded audio-visual *localisation*, similarly to a previous study with infants (Neil *et al.*, 2006) but in contrast with studies in which older children responded to a central target (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011), we used a task in which participants judged whether an audio- and/or visual target was presented to their left or right. To evaluate whether performance could be explained by parallel processing of signals, we compared distributions of response times with predictions of the classic race model (Miller, 1982), a recently formulated extended race model (Otto & Mamassian, 2012), and a pooling ("coactivation") model.

Methods

Participants

Participants were aged 4 (mean (SD) age = 4.7 (0.3) years, n = 17, 7 male), 5 to 6 (mean (SD) age = 6.1 (0.6) years, n = 21, 11 male), 7 to 8 mean (SD) age = 8.0 (0.6) years, n = 32, 18 male), 9 to 11 (mean (SD) age = 9.9 (0.5) years, n = 25, 12 male), or were adult (mean (SD) age = 25.6 (5.0) years, n = 26, 11 male). An additional 6 participants (5 aged 4 years, one aged 5 years) who had fewer than 24 valid trials in any condition were excluded (see *Analysis*). Participants were recruited from a database of volunteers and from schools in London. The study protocol was in accordance with the Declaration of Helsinki and was approved by the local ethics committee. Participants were tested with parents' or their own informed consent.

Stimuli and procedure

Participants pressed buttons to indicate whether stimuli appeared to the left or to the right of fixation. There were four trial types: those in which stimuli were an auditory (*A*) beep, a visual (*V*) flash, a simultaneous auditory-visual (*AV*) beep and flash on the same side; or blank *(B)* trials in which there was no stimulus. Stimuli appeared at unpredictable intervals of 500 – 1000 ms after the fixation cross disappeared. The purpose of the blank trials, together with the variable interval, was to make the stimulus onsets less predictable. Stimuli were presented on a monitor subtending 39.4° visual angle from a distance of 65 cm while using a chin rest. Fixation was monitored using a remote eye tracker (EyeLink 1000, SR Research, Mississauga, Ontario, Canada). Stimuli were presented using the Psychophysics Toolbox for Matlab (Brainard, 1997). Testing was carried out in a quiet space with low ambient light. The monitor had a background luminance of 0.5 cd/m^2. Each trial began with participants fixating a central 1° cross. After 300 ms in which eye position, as

determined by the eye tracker, remained within a 2.4° radius of the centre of the cross, the cross disappeared. After a randomly chosen interval between 500 - 1000 ms (in steps of 100 ms), during which participants maintained eye position within the central 2.4° region, one of the four trial types (*A*, *V*, *AV* or *B*) took place. If participants moved their eye position out of the central region during the interval, the fixation cross was shown again and the trial was restarted.

Participants completed up to 8 blocks, each block containing 6 trials each of the four different types randomly mixed, with stimuli appearing equally often on the left or right within each block and stimulus type. On auditory (*A*) trials, a 100 ms 2 kHz 84 dB SPL pure tone was played by a speaker mounted at the edge of the screen, 26.9° to the left or right of the screen centre. On visual (*V*) trials, a circular 100 ms 1° 13.9 cd/m^2 flash was presented 20.3° to the left or right of the screen centre. On audio-visual (AV) trials, both the sound and the flash were presented simultaneously on the same side of the screen. Auditory and visual intensities were at values established during piloting to elicit similar mean response times. To achieve this, the visual stimulus had to be made relatively low-contrast and the auditory stimulus relatively loud. The small (6.6°) offset between the targets' auditory and visual locations was due to the experimental setup, in which sounds were delivered by loudspeakers fixed to the sides of the monitor.

Participants had the forefinger of each hand positioned over a response button, located in front of them on their left and right at a comfortable distance. They were instructed to press a button as quickly as possible to indicate the side on which each target appeared. On a correct press of the button the target was replaced with an image of one of a set of farmyard animals. For younger children, these animals formed the basis for a secondary game whose purpose was to maintain their interest in the task. Children kept track of how many times a particular animal appeared and, after a set number of times, the experimenter moved a counter on a board game. The

children could win stickers on the basis of the board game. All data collection was carried out by an experimenter expert in testing children. To keep the test feasible with children as young as 4 years, participants were permitted to take breaks from being on the chin-rest, for example to look at the board game, or to take a break when they wished. Participants were also allowed to stop before the end of 8 blocks if they wished. After each trial was complete, the experimenter pressed a key to initiate the next trial. This allowed the experimenter to maintain an appropriate pace for participants of all ages, including, for younger participants, time for the board game or breaks as needed. The next trial began with a fixation cross as described above. The study was preceded by a block of 5 practice trials, which was repeated if necessary.

Analysis of response times

Correct responses at times of up to 2500 ms were retained for analysis. Within this range, all groups except adults showed a subset of early responses at very low latencies that are assumed to be false alarms unrelated to the stimulus. Such responses are normally excluded by use of a criterion, e.g. 100 ms in Otto & Mamassian (2012). In our study, which included age groups with very different response time distributions, setting a single criterion across all age groups would be unlikely to capture the "early" cut-off correctly at all ages. To ensure that comparisons of the main distribution across age groups were valid, and that all ages exclude the "early" outliers that cannot be well fitted by the models as part of the main distribution of reaction times, we set an individual criterion for each age group by identifying a break-point between the normal distribution of reaction times and early outliers. Based on this, 3.4% or fewer trials were excluded at any age. See Supplemental Methods and Supplemental Fig. S1 for full details. Participants who had at least 24 correct trials remaining in the (lower criterion - 2500 ms) range in all three stimulus conditions were included in the analysis. Compared with the maximum possible number of valid trials over

the whole experiment, 144, the mean (SD) numbers of trials of those included in the analysis were 119.9 (10.3) at age 4 , 127.8 (9.6) at 5-6, 132.2 (6.4) at 7-8, 137.5 (4.6) at 9-11, and 141.4 (2.1) in adults. Descriptive statistics (mean and SD of response times) were calculated for each subject and condition, and means of these were calculated and compared using ANOVA and t-tests.

Model predictions

A crucial question is whether any speed gains when responding to two signals given together *vs.* either alone can be explained by parallel processing, or whether they indicate that signals are being combined (pooled). To answer this, we compared empirical distributions of *AV* response times with the predictions of three models: the classic race model (Miller, 1982), an extended race model (Otto & Mamassian, 2012), and a pooling (coactivation) model. These models aim to predict the *AV* distribution based on the separately measured *A* and *V* distributions, together with assumptions about the information processing operations that underlie response times. The three models are shown schematically in Fig. 1a-c and described in outline below; full details are given in the Supplemental Methods.

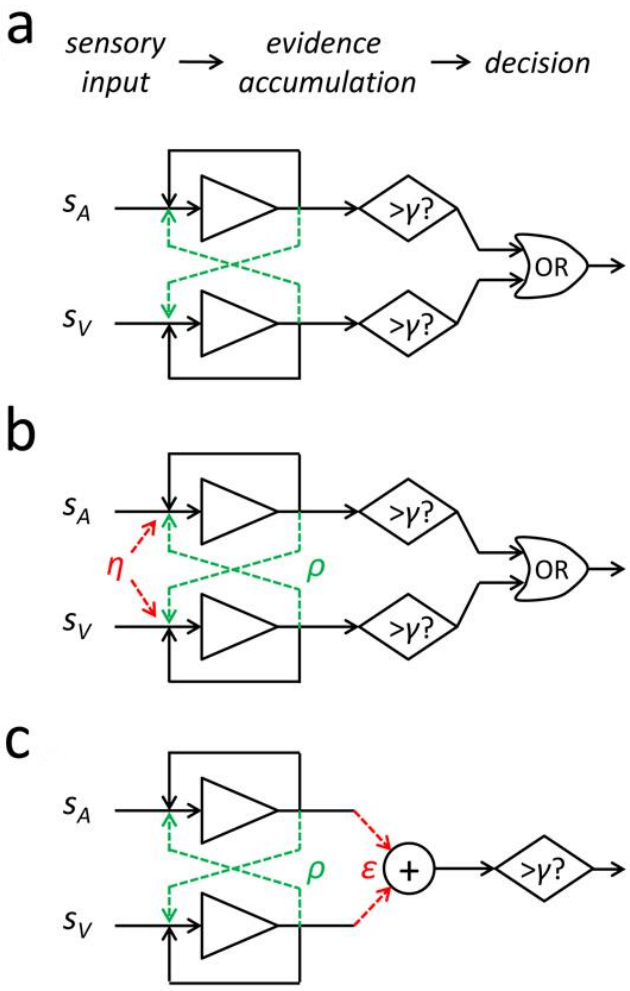

Figure 1. a. Models predict how incoming sensory evidence is accumulated until a decision is made. Here, auditory and visual sensory evidence for the presence of a stimulus (S_A, S_V) is accumulated until it passes a threshold (γ), which triggers a response. In Miller's (1982) model, each channel accumulates evidence independently, and decisions are made via a race between (logical "OR" operation on) their outputs. The model assumes an interaction between channels (dotted lines) leading to a negative correlation between their distributions of response times (RTs) as well as context invariance (see text). **b.** The extended race model (Otto & Mamassian, 2012), does not assume context invariance, but allows that processing two stimuli at once may add variability (random noise) to accumulation times, described by a free parameter η. A second parameter ρ describes the between-channel interaction. **c.** In the pooling (coactivation) model, decisions are made based on evidence pooled from both channels. A correlation parameter ρ plays the same role as in (c). An efficiency parameter (ε) describes the efficiency of pooling when two inputs must be accumulated at once.

Classic race model

Miller's (1982) classic formulation of the race model has been widely used to analyse response times in previous multisensory studies with adults (*e.g.* Schroger & Widmann, 1998; Hughes *et al.*, 1994) and children (Neil *et al.*, 2006; Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011; Barutchu *et al.*, 2011); Fig. 1a. Model predictions for combined-cue condition *AV* are calculated from cumulative distributions of *A* and *V* response times. The model prediction ("Miller's bound"; see Supplemental Methods for full details) is given simply by the sum of the cumulative *A* and *V* distributions. Importantly, the model does not aim to capture performance, but rather to state the upper limit on the speed gains theoretically possible via a race between channels. To do so, the model assumes (1) negative correlation between channels (see Fig. 1a), an extreme case predicting the greatest RT benefits (see Supplemental Methods), and (2) "context invariance" (Luce, 1986): that the system's processing of one signal (*e.g. A*) is the same when *A* is the only signal present and when a second signal is also present (*e.g. AV*). If no further assumptions are made, then response times faster than this bound cannot be explained by parallel processing. Times faster than this prediction have traditionally been interpreted as showing evidence for pooling of signals. In the present study, we used this approach to compare *AV* reaction times with Miller's bound prediction.

Extended race model

It is possible that processing one signal may come at some cost to processing the other. In that case, the "context invariance" assumption embodied in Miller's classic (1982) model does not hold. In a recent extension of Miller's race model, Otto and Mamassian (2012) argued that Miller's bound could be exceeded at low quantiles if parallel processing of signals made response times more variable than in single-cue conditions. Increased variability would make the fastest times faster and the slowest times slower than model predictions. Indeed, this is the observed pattern in

previous data sets, including those from developmental visual-auditory studies (Neil et al., 2006, Fig. 3b; Barutchu et al., 2009, Figs. 3b, 3c, 3d; Barutchu et al., 2010, Fig. 3; Brandwein et al., 2011, Fig. 2). However, the violation only at lower quantiles may also be explained by consideration of only the extreme case of total negative correlation. Otto and Mamassian developed an extended race model predicting two-cue response time distributions from single-cue distributions with the degree of correlation as a free parameter (ρ; see Fig. 1b). They also did not assume context invariance, but that processing two signals in parallel could add variability to response times; this added variability was a second free parameter (η). When η = 0 and ρ = -1, the model is equivalent to Miller's model. They showed that for sensory tasks with arbitrary pairings of stimuli (motion and tone, motion and colour), a classic race model analysis showed violations of Miller's bound. This result would previously have been considered to provide good evidence for pooling of signals. However, they were able to account for the distribution of responses by parallel processing in which signals were moderately negatively correlated, and parallel processing of two signals added some variability to response times. This analysis poses a significant challenge to previous results that have used violations of Miller's bound as evidence for pooling of sensory signals (*e.g.* Schroger & Widmann, 1998; Hughes *et al.*, 1994), including all the developmental audio-visual studies to date (Neil *et al.*, 2006; Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011; Barutchu *et al.*, 2011).

Accordingly, in the present study we also fitted the extended race model with free parameters for correlation (ρ) and noise (η) to each participant's response time distributions. We used the approach described in Otto & Mamassian, 2012; see Supplemental Methods and the article for full details. The approach relies on fitting normal distributions to reciprocal response times (*i.e.* 1/RT, which are response *rates*; the fitted distributions are "recinormal distributions"). Model predictions are made in the 1/RT domain. Recinormal distributions typically provide good fits to

reaction time data, and can be interpreted as reflecting normally distributed variability in the drift of an evidence accumulation mechanism (see Otto & Mamassian, 2012; Reddi, Asrress, & Carpenter, 2003; Carpenter & Williams, 1995).

Pooling (coactivation) model

It has often been assumed that reaction time benefits exceeding those predicted by race (parallel channels) models show evidence for "coactivation", or the pooling of sensory signals. In the audiovisual domain, conclusions about pooling based on inabilities of race models to fit the data have been the standard approach (*e.g.* in Hughes *et al.*, 1994; Schroger & Widmann, 1998; Neil *et al.*, 2006; Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011; Barutchu *et al.*, 2011). Clearly, the strongest conclusions about the processing of sensory signals would come from a direct comparison of race and coactivation models. In the case of a model with free parameters, like the extended race model (Otto & Mamassian, 2012), some degree of fit to the data is expected. In that case, it would be difficult to make reliable conclusions about pooling without an explicit comparison to the fit provided by an alternative pooling model. For this reason we also formulated a pooling (coactivation) model.

There have been detailed studies describing theoretical properties of information processing systems using single or combined channels, particularly by Townsend and colleagues. These have included the formulation of reaction time predictions for "coactivation" as well as for parallel channels (*e.g.* Grice, Canham, & Boroughs, 1984; Miller, 1986; Schwarz, 1989; Diederich, 1995; Townsend & Nozawa, 1995; Townsend & Wenger, 2004; Townsend & Eidels, 2011). For our analysis, we formulated a coactivation model which combines the approaches of Otto and Mamassian (2012) and of Townsend and colleagues (Townsend & Wenger, 2004; Townsend & Eidels, 2011); see Fig. 1c.

Above, we described how a race between parallel signals can in theory lead to reaction time improvements relative to those based on a single signal. Here we describe how pooling of signals can also lead to reaction time improvements. We then introduce the coactivation model that we used to analyse our data.

Sensory decision-making can be understood to depend on the accumulation over time of evidence in favour of a hypothesis (for reviews, see Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Gold & Shadlen, 2007). If multiple independent sources of evidence for a hypothesis are available (*e.g.* both auditory and visual evidence that a stimulus is present), then evidence will be accumulated at a faster rate than when only one source is available, and will therefore (given a constant decision bound) produce faster reaction times. An increase in the rate of evidence accumulation also leads to less variable reaction times[1]. Pooling two estimates rather than relying just on one would therefore be expected to result in faster accumulation of sensory evidence, and faster and less variable reaction times.

Our coactivation model (Fig. 1c) predicts cumulative distributions of two-cue response times for a participant responding based on pooled evidence from two redundant sensory signals. Model predictions for the distribution of *AV* rates are made based on *A* and *V* distributions and two free parameters; correlation between signals (ρ) and the efficiency of pooling (ε). As in the extended race model, predictions are made in the $1/t$ domain. Reciprocal response times (*i.e.* response *rates*) are modelled as coming from a normal distribution whose mean is the mean rate of responding, and whose variance is the variance in this rate. The rate can be understood to be proportionate to the rate at which evidence is accumulated towards a decision bound (Bogacz *et*

[1] This is because the rate and the variance in this rate are defined in units of *1/time*, and increases to the rate alone in units of *1/time* lead to reductions in both the mean and the variance in *time* units (*e.g.* seconds). See also the description of recinormal distributions above in section *2. Extended race model*, and in the Supplementary Methods.

al., 2006; Gold & Shadlen, 2007). The special case of perfect efficiency ($\varepsilon = 1$) is equivalent to context invariance, but as in the extended race model context invariance is not assumed. Full details are in the Supplemental Methods.

Results

Accuracy

Mean percentages of correct left / right responses (within the range of times between each group's "early" cut-off and 2500ms, and for participants who provided enough valid trials in each condition to be in the main analysis) were 94.7% at 4, 97.2% at 5-6, 97.3% at 7-8, 98.6% at 9-10 years, and 99.1% in adults. In other words, while the youngest children made occasional errors, older children and adults almost never did.

Response times

We first analysed changes in mean and standard deviation (SD) of correct response times by age and condition. Overall, both the mean (Fig. 2a) and the standard deviation (Fig. 2b) of response times decreased with age. An ANOVA of mean response times (Fig. 2a) revealed that there were main effects of *Stimulus*, $F(2, 348) = 53.5$, $p<0.001$ and of *Age*, $F(4, 348) = 80.0$, $p<0.001$, but no interaction, $F(8, 348) = 0.5$, $p = 0.86$. For mean SDs of response times (Fig. 2b), there were main effects of *Stimulus*, $F(2, 348) = 315.7$, $p<0.001$, of *Age*, $F(4, 348) = 49.5$, $p<0.001$, and a *Stimulus x Age* interaction, $F(8, 348) = 6.9$, $p<0.001$. Thus, performance improved with age, and varied across

conditions. While for mean times these effects were additive, for variability (SD) of times there was an interaction showing changes in variability across conditions with age (Fig. 2b).

Planned comparisons (paired t-tests) were used to compare *AV* measures at each age with both *A* and *V* measures. Significantly better performance in the combined-cue condition *AV* than in *both* single cue conditions *A* and *V* would be evidence that participants were exploiting parallel audio-visual information to make faster (Fig. 2a) or less variable (Fig. 2b) decisions than with the best single cue. At all ages, both means and SDs of response times were significantly lower in the *AV* condition than in both the *A* and the *V* conditions (Fig. 2a-b, *s, and Supplemental Table S1). Thus, children as young as 4 years exploited parallel audio-visual information to respond faster, and less variably, than when given either information source alone.

The absolute reduction in response time when two cues were available was similar across age groups, as seen in Fig. 2a and in the non-significant *Age* x *Stimulus* interaction. Because unimodal latencies decreased with age, when expressed as a proportion of unimodal response time (as a relative multisensory enhancement, rMRE), the two-cue advantage shows an increase in age. Calculating rMRE for each participant as (min(A|V) - AV) / min(A|V) * 100%, the mean (SEM) effect by age group (in %) is 12.5 (1.8), 17.1 (1.9), 19.1 (1.0), 20.7 (2.2), and 25.0 (1.0).

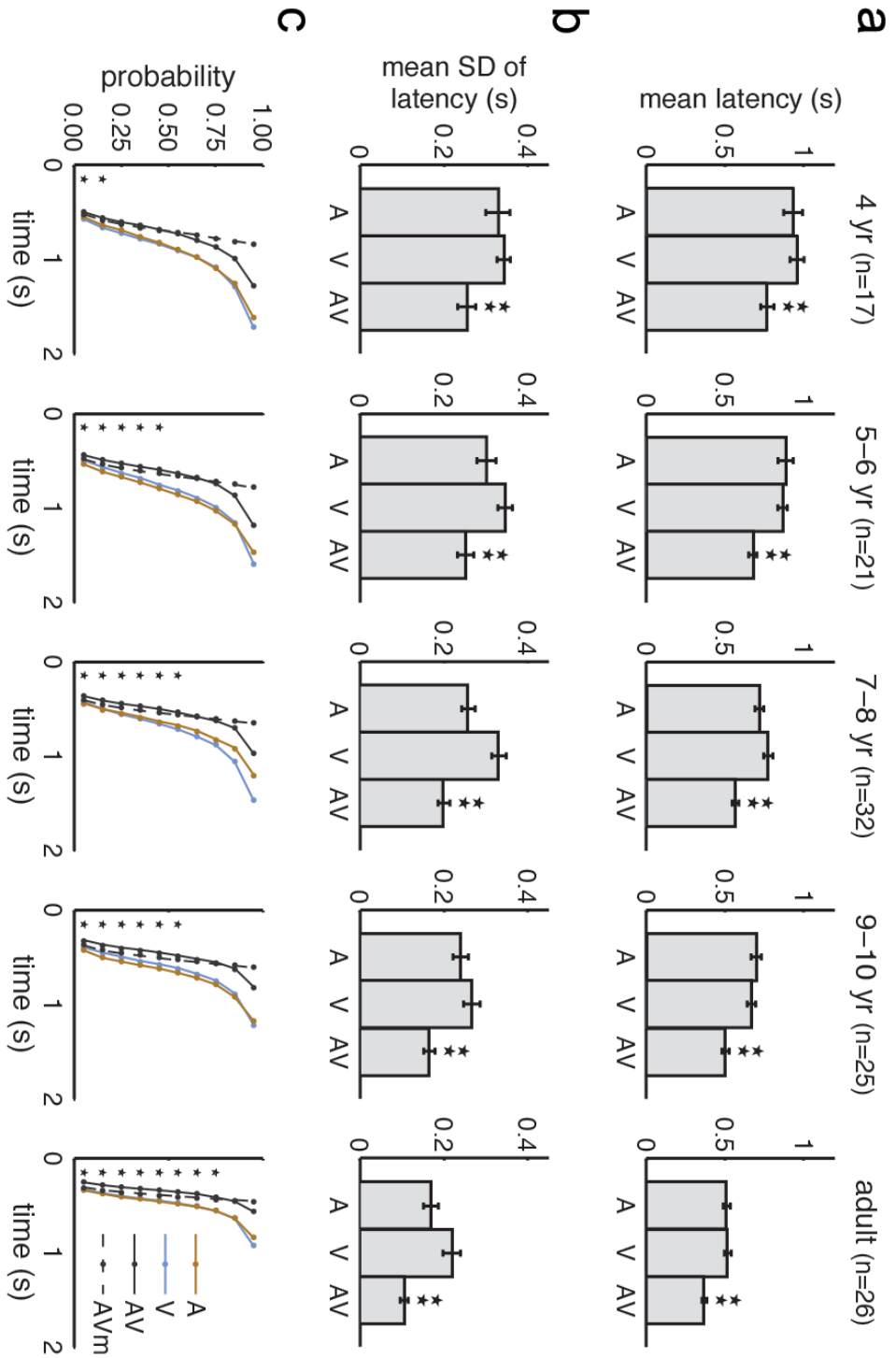
a
b
c
4 yr (n=17)
5–6 yr (n=21)
7–8 yr (n=32)
9–10 yr (n=25)
adult (n=26)
mean latency (s)
mean SD of latency (s)
probability
time (s)
A
V
AV
AVm

Figure 2 a. mean ± SEM of the mean latency and **b.** mean ± SEM of the SD of latency for each participant. ** *AV* differs from *A* and *AV* differs from *V* at the 5% level on paired two-tailed t-tests. **c.** Mean cumulative distributions of times in *A*, *V*, and *AV* conditions, compared with the classic race model (*AVm*), i.e. Miller's bound (Miller, 1982), at 10 probability deciles. * mean AV time significantly faster than *AVm* prediction at the 5% level (Šidák-corrected for multiple comparisons) on paired one-tailed t-test.

Classic race model

Some multisensory speed gains can be predicted by parallel processing. As a first test of the extent to which parallel processing could explain our empirical data (Fig. 2a), we compared distributions of two-cue (*AV*) response times with those predicted by Miller's bound - that is, the fastest possible times given parallel processing of the single cues as measured in conditions *A* and *V* under the assumptions of Miller's (1982) model. We compared mean measured *AV* latencies with mean model predictions (Fig. 2c, *AV vs. AVm*) at each of 10 probability deciles from 0.05 to 0.95 using one-tailed paired t-tests, Šidák-corrected for multiple comparisons. At all ages, at least two deciles showed speeds significantly exceeding Miller's bound (Fig. 2c, *, and Supplemental Table S2). The number of deciles exceeding the bound increased from 2 of 10 at 4 years to 8 of 10 in adults (Fig. 2c). The significant speed gains relative to Miller's bound indicate that children as young as 4 years on average performed faster than could be explained by Miller's model. This stands in contrast with several previous studies in which only children at older ages showed this result (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011), At all ages in the present data, the lowest deciles are those at which violations of the model occur, a pattern commonly observed in other tests of Miller's bound (e.g. Brandwein *et al.*, 2011; Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Neil *et al.*, 2006).

Extended race model

It has been pointed out that some speed gains relative to the Miller bound may be explained by increased variability when processing single cues, and so need not necessarily indicate coactivation (Otto & Mamassian, 2012). To test whether this could explain our data, we compared distributions of two-cue (*AV*) response times with those predicted by the extended race model developed by Otto and Mammasian (2012). Testing the model involves, first, fitting Gaussian

distributions to reciprocal response times in conditions *A*, *V*, and *AV*. The mean square errors of these empirical fits are plotted in Fig. 3, and the best-fitting parameters are given in Supplemental Table S3. Next, these distributions fitted to conditions *A* and *V* are used, together with free parameters for correlation (ρ) and noise (η), to fit a model prediction to condition *AV*. The error of the extended race model fit ("*r*") is also shown in Fig. 3. As might be expected, approximations of the empirical *A*, *V*, and *AV* distributions by recinormal distributions include some error. This need not denote systematic mis-fits; it is also expected due to sampling noise. Fitting of the extended race model prediction *"r"* to empirical *AV* data also includes error (Fig. 3). As Fig. 3 shows, extended race model predictions for *AV* distributions (*"r"*) have higher error than the recinormal fits to empirical *A*, *V*, and *AV* distributions, and this error increases with age. This implies that besides noise due to sampling, the model fit may include a systematic mis-fit, increasing with age.

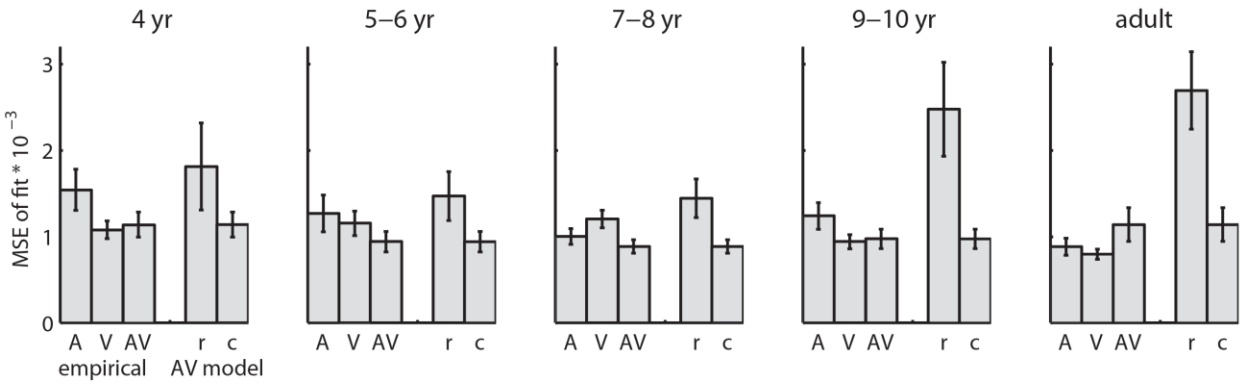

Figure 3. Mean ± SEM mean square error (MSE) of cumulative recinormal fits to empirical response times in conditions *A*, *V*, *AV*, and of extended race model (*r*) and coactivation model (*c*) fits to empirical response times in condition *AV*.

Fig. 4a shows cumulative response time distributions of representative individuals of different ages for empirical *A*, *V* and *AV* times (points and solid lines), and the best-fitting extended race model-predicted distribution for *AV*, *AVm* (red dashed line). While at younger ages deviations from the model tend to be small (red dashed line vs. solid black line), at older ages the empirical data deviate substantially and systematically from model predictions. However, even in older ages differences are not readily apparent from these plots (which also show only a single example participant from each group). To draw conclusions across all participants we carried out an analysis of both central tendency and dispersion, asking to what extent empirical distributions differed from those predicted by the model for each individual in the study. We then made a statistical comparison of empirical and model-predicted values for each age group.

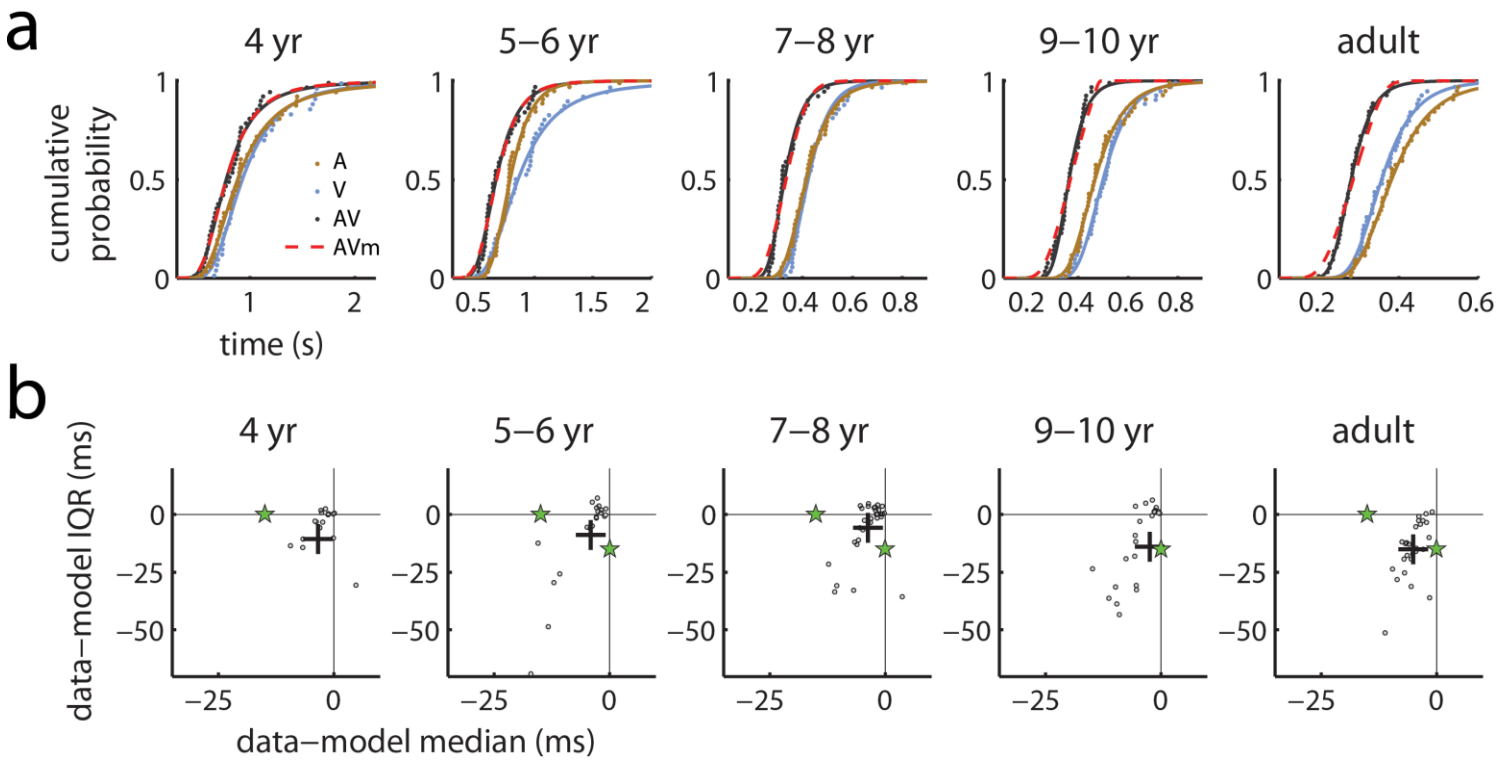
a
4 yr
5–6 yr
7–8 yr
9–10 yr
adult
cumulative probability
A
V
AV
AVm
time (s)
b
data–model IQR (ms)
data–model median (ms)

Figure 4. Extended race model analysis. **a.** Data from representative individuals, one at each age: empirical reaction times (data points), fitted recinormal cumulative distributions (solid lines), and best-fitting extended race model prediction for AV times (dashed red line) for one representative participant from each age group. **b.** Group data showing differences between medians and inter-quartile ranges (IQRs) of recinormal fits to cumulative distributions of empirical AV response times and those predicted by the model via least-squares fits to cumulative AV response time distributions, for individual participants (data points) and as a group mean (+). * group mean differs significantly from zero on two-tailed one sample t-test (see Supplemental Table S4).

A potential concern when empirical data are fitted by distributions (here, reciprocal reaction times are fitted with Gaussian distributions) is that if empirical distributions systematically differ from those fitted, the fitting process could lead to systematic biases in the estimation of parameters from the data. To check this, we compared model-fitted and empirical measures of centre (mean, median) and dispersion (standard deviation, SD; interquartile range, IQR). We found that the approach of fitting Gaussian distributions to reciprocal response times produced biased estimates of mean and SD in our data, while median and interquartile range (IQR) were more robust against such biases; see Supplemental Fig. S2. We therefore used median and IQR as summary measures of centre and dispersion. We analysed the differences between median and IQR of fits to empirical *AV* distributions, and median and IQR of extended race model predictions for the *AV* condition. Fig. 4b plots these differences for each individual subject, and the group mean difference (+ symbol). Two-tailed one-sample t-tests tested the hypothesis that each group's mean median and IQR difference from the model differed from zero (Fig. 4b, *, and Supplemental Table S4). At all ages at least one of these two measures showed performance to be better than the model, and at three ages (5-6, 7-8, adult), both measures were significantly better. Thus, on average, the best-fitting extended race model using two free parameters per observer failed to capture the gains in speed and/or variability of response times from 4 years upwards. The deviations from the model were small but highly systematic.

In summary, from 4 years, response times to simultaneous visual and auditory information were significantly faster and/or less variable than those predicted by the extended race model with ρ and η parameters providing the best fit to cumulative response time distributions.

Although the extended race model did not correctly predict distributions of reaction times, Table 1 reports the mean best-fitting values for ρ and η by age group, for comparison with previous

results (Otto & Mamassian, 2012) or those from future studies. These fits indicate that the age changes in *AV* performance depend on a progressively greater negative correlation between signals and on progressively increasing noise. ANOVAs found that age changes in both parameters were significant: for ρ (correlation), $F(4,116) = 10.0$, $p < 0.001$; for η (added noise), $F(4,116) = 17.7$, $p < 0.001$. A conclusion of much greater noise in the mature as compared with the developing system ($\eta = 1.04$ in adults *vs* 0.30 at 4 years; Table 1) would certainly be counter-intuitive. Together with the findings that this model's best fitting parameters do not give a good account of the empirical data (Fig. 4b), this argues against the extended race model as a good explanation of the empirical results.

Table 1. Mean (SEM) for parameters ρ and η in the extended race model that provided least-squares fits to distributions of *AV* response times in each age group.

age group	*ρ (correlation)*	*η (added noise)*
4 yr	0.06 (0.14)	0.30 (0.05)
5-6 yr	-0.33 (0.12)	0.47 (0.06)
7-8 yr	-0.42 (0.08)	0.51 (0.05)
9-11 yr	-0.51 (0.12)	0.67 (0.07)
Adult	-0.86 (0.05)	1.04 (0.07)

Coactivation model

We next compared distributions of two-cue (*AV*) response times with those predicted by our coactivation model. The first stage of this, fitting Gaussian distributions to reciprocal response times in conditions *A*, *V*, and *AV*, is the same as for the extended race model. Fig. 3 plots the errors of these fits. At the next stage, the distributions fitted to conditions *A* and *V* are used, together with free parameters for correlation (ρ) and efficiency (ε), to fit a model prediction to

condition *AV*. Fig. 3 plots the error of this model fit ("*c*"), alongside the error of the extended race model fit ("*r*"). It is clear from Fig. 3 that for condition *AV*, the error in coactivation model-predicted *vs.* empirical reaction times (*c*) is the same as the error in fitting a recinormal distribution to describe empirical AV times (*AV*). Indeed, for all individual participants in all age groups, the recinormal and model-fitted parameters were identical. Likewise, model-predicted median and IQR values exactly matched those of recinormal-fitted *AV* distributions. To be precise, in all participants these values differed by $<10^{-4}$ seconds, reflecting the tolerance values in our search for parameters that minimise least-squares[2].

The coactivation model is able, in principle, to match recinormal-fitted *AV* data exactly because, being based on a sum of two normally distributed variables (see Supplemental Methods, Eqs 5-6), it always predicts a normal distribution of reciprocal reaction times. Note, however, that being able to fit all observers' data meant that for all observers, correlation and efficiency parameters within permitted bounds could be found; that is, all observers could be fitted with a correlation ranging from -1 to +1 and an efficiency ranging from 0 to 1 (the fitted parameters are described in more detail below).

Paired t-tests were used to compare MSEs of fits from the extended race model (Fig. 3, "*r*" *vs.* "*c*"). These tests excluded one extreme outlier, a 4-year-old with a very high race model MSE, for whom the difference between race and coactivation model MSEs was more than 10 times the interquartile range greater than the median of these differences in this age group. As

[2] We used the Matlab function *fminsearch* with tolerance parameters *TolFun* and *TolX* set to their default values of 10^{-4}. This meant that the search ended when either the MSE or the parameter values changed by less than 10^{-4} on two consecutive iterations. These tolerance values do not directly predict how far medians or IQRs of fitted reaction times might diverge from one fit to another, but they do predict that this divergence should be low, but non-zero.

Supplemental Table S5 reports, MSEs were significantly lower for the coactivation model at all ages.

Table 2 reports the means for the best-fitting parameters ρ (correlation) and ε (efficiency) by age group. These means indicate that correlation changed from being somewhat more negative (-0.45 at 4 years) to somewhat less negative (-0.33 in adults). At the same time, efficiency improved, from 0.60 at 4 years to 0.68 in adults. ANOVAs found that age changes in ε (efficiency) were significant, $F(4,116) = 8.1$, $p < 0.001$, while age changes in ρ (correlation) were not; $F(4,116) = 1.1$, $p = 0.34$. A conclusion of age improvements in efficiency – here denoting the proportion of potentially useful signal used when pooling evidence from two sensory channels towards a single decision – is plausible.

Table 2. Mean (SEM) for parameters ρ and ε *(efficiency)* in the coactivation model that provided least-squares fits to distributions of *AV* response times in each age group.

age group	*ρ (correlation)*	*ε (efficiency)*
4 yr	-0.45 (0.046)	0.60 (0.009)
5-6 yr	-0.45 (0.045)	0.63 (0.012)
7-8 yr	-0.44 (0.044)	0.64 (0.005)
9-11 yr	-0.41 (0.052)	0.66 (0.013)
Adult	-0.33 (0.041)	0.68 (0.009)

Discussion

We used a speeded audiovisual localisation task to investigate the development of children's abilities to pool multisensory information. As early as four years, the youngest age we tested, children were faster and less variable in their response times given both sensory signals compared with either one alone (Fig. 2a-b). These advantages were greater than those predicted by a classic race model widely used to analyse response time distributions (Fig. 2c). They were also greater than those predicted by an extended race model that did not assume "context invariance", but allowed for increased variability when processing two signals (Fig. 4b). However, the multisensory advantages at all ages were predicted by a coactivation model predicting responses based on signals that are pooled ahead of a decision. The deviation from race model predictions and consistency with coactivation (pooling) was seen across the age range 4-12 years and in adults, showing early developing and sustained abilities to pool auditory and visual signals for speeded spatial localisation. Alongside this overall effect, there were major age-related reductions in overall speed and variability of response times, and in the efficiency of pooling as estimated by the coactivation model.

The overall age improvements in response times and their variability are likely to reflect, firstly, improvements in the speed and spatial acuity of visual and auditory spatial estimates, and secondly, improvements in abilities to map the left/right manual response onto the left/right spatial signal. To keep children motivated, they participated in a secondary game in which they kept track of how many times a particular animal appeared following responses at the target location. It is possible that this additional demand provided some dual-task disadvantage for children compared with adults, in which case some of the age differences found may also be attributable to this demand.

Importantly, the study showed no evidence for a qualitative change in the signal processing strategy over the age range. At all ages, behaviour departed from predictions of parallel sensory decision making and were in line with the predictions of pooling of sensory signals ahead of a decision. Nevertheless, the extent to which behaviour departed from race models increased with age, suggesting extended development of sensory integration through childhood. At one extreme, adults showed an average 25% speed gain, violations of Miller's bound at nearly all (8 of 10) quantiles (Fig. 2c), medians and IQRs of response times better than predicted by the extended race model for most individuals (Fig. 4b), and a mean estimated pooling efficiency of 0.68 (Table 3). At the other extreme, four-year-olds showed an average 12% speed gain, violations of Miller's bound at only two of 10 quantiles, medians and IQRs of response times similar to extended race model predictions for many individuals (Fig. 4b), and a mean estimated pooling efficiency of 0.60 (Table 3).

Our analysis suggests that the increasing departure with age from the predictions of either race model is explained by the increasing efficiency of pooling. At younger ages, pooling is less efficient (more potentially useful information is lost), so the benefits of pooling are smaller, and behaviour departs less dramatically from the range that can be predicted by parallel-channels models. Age improvements in pooling efficiency would be consistent with childhood brain changes including synaptic pruning and myelination (Huttenlocher & Dabholkar, 1997; Giedd *et al.*, 1999), although their dependence on development of specific cortical or subcortical networks remains to be investigated. Another factor contributing to stronger departures from the model in older child and adult observers' data may be the somewhat greater number of data points for adults, or a more consistent (less noisy) patterns of responding. Finally, although instructed to respond as quickly as possible, all observers presumably used some speed/accuracy trade-off, the setting of which is unknown and could have varied across age groups.

These results can reconcile those from previous studies in which infants showed speed gains exceeding race models in a localisation task (Neil *et al.*, 2006), but children aged below 7 or 10-12 years did not do so reliably in a central task where the stimulus was always at the same location, straight ahead (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011). In the present left/right localisation task, we saw evidence for pooling of signals from 4 years. One might expect that the additional need to make correct spatial decisions would work against finding the greatest speed improvements. However, it may be that the need to localise the stimulus in space on every trial tapped into multisensory decision mechanisms dedicated to spatial localisation. A well-known substrate for this kind of localisation is the superior colliculus (SC; Wallace *et al.*, 1996; Wallace & Stein, 1997; Stein & Meredith, 1993). Multisensory integration in superior colliculus develops with postnatal sensory experience, which seems to be learnt via the cortico-SC projection (Stein *et al.*, 2009). The decisions in our study may also (or alternatively) have depended on other cortical mechanisms (Molholm *et al.*, 2002; Martuzzi *et al.*, 2007; Noppeney *et al.*, 2010). A study with adults directly comparing saccadic, directed manual and simple manual responses found violations of Miller's bound in all three cases (Hughes *et al.*, 1994). The present developmental results from a directed manual task, together with previous data from spatial orienting (Neil *et al.*, 2006) and simple manual tasks (Barutchu *et al.*, 2009; Barutchu *et al.*, 2010; Brandwein *et al.*, 2011), suggest that pooling of auditory and visual signals may develop much earlier for spatial localisation of signals than for simple detection of their onset. A follow-up study comparing simple responses and localisation directly should test this interpretation.

Together, the good predictions of empirical reaction time distributions, the lower MSEs of fits, and the more interpretable age-changes in model parameters give strong reason to favour the coactivation model. Of course, the race and coactivation models we tested represent one variant of each, and it is possible that alternative models of either race or coactivation mechanisms would

behave differently. Nevertheless, these results are the first to meet the challenge posed by a recent critique of standard analyses of multisensory decision making (Otto & Mamassian, 2012). In that study, violations of the race model for judgments based on arbitrary pairings of sensory stimuli (*e.g.* motion direction and colour change) could be shown to be explicable by parallel decision making. In our study, spatial location judgments based on visual and auditory signals - a common and ecologically important function - was not explicable in this way. This is not surprising given the evidence for numerous cortical and sub-cortical mechanisms for pooling of visual and auditory information about target onsets (*e.g.* Wallace *et al.*, 1996; Wallace & Stein, 1997; Stein *et al.*, 2009 Molholm *et al.*, 2002; Martuzzi *et al.*, 2007; Noppeney *et al.*, 2010). In a future study it would be ideal to make direct comparisons to test how differences in model fit are related to the choice of stimulus pairing.

While we found evidence for departures from race models, and consistency with coactivation, in our youngest group, 4 years, the earliest age at which this behaviour would be evident is not yet known. There has been evidence from infants, using a reflexive orienting task and a classic race model analysis, for speed gains exceeding the Miller race model at 8-10 months (Neil *et al.*, 2006). To understand development below 4 years, future studies should also test the extended race and coactivation models used here, and should evaluate potential differences between reflexive and voluntary tasks.

It is interesting to compare our evidence for multisensory speed gains *via* pooling of signals from at least 4 years of age stands with results from numerous developmental studies measuring multisensory precision, in tasks that were not speeded. These tasks have found absent or sub-optimal pooling of signals until much later in childhood (Gori *et al.*, 2008; Nardini *et al.*, 2008; Nardini *et al.*, 2010; Gori *et al.*, 2012; Nardini *et al.*, 2013; Petrini *et al.*, 2014). In these studies,

children aged 8 years or older did not improve their perceptual discrimination thresholds by taking correctly weighted averages of multiple estimates (*e.g.*, in Gori *et al.*, 2008, visual and haptic estimates of the heights of blocks). This highlights the gap, also in the adult literature, between research on the speed and the accuracy of multisensory decisions. Real-world behaviour requires the constant setting of speed/accuracy trade-offs. The present work provides a basis for future studies to model both the time course and the precision of decisions in developing sensory systems. The development of an "ideal observer" pooling model in the present study may be a useful step towards this. Interestingly, our model suggested that even adults showed only 68% of their theoretically optimal pooling efficiency in terms of decision speed. In contrast, adults making perceptual judgments that are not time-limited often show near-100% efficient (or near-Bayes-optimal) performance (e.g. Ernst & Banks, 2002; Alais & Burr, 2004). An intriguing question is to what extent this represents a speed-accuracy trade-off, and whether a single model could capture the efficiency (relative to an ideal observer) of both the speed and the accuracy of decisions, including shifts in these as task demands are changed. Introducing rewards that depend on both time and accuracy can result in "ideal observer" predictions for the speed / accuracy trade-off itself. Another question is in what circumstances observers can find optimal trade-offs (*i.e.* those that maximise their reward rate; Gold & Shadlen, 2002), and how these abilities develop.

The greatest challenge for explaining sensory and multisensory development – in the present case, two-fold improvements in unisensory speed, together with improving efficiency in pooling – is to understand the mechanisms of change. Such mechanisms need to be understood at several distinct levels of analysis – for example, following Marr's (1982) influential formulation, both the algorithms used for sensory processing and their "biophysical" underpinnings need to be understood, which can be considered to provide separate (albeit inter-related) levels of

explanation. An important challenge, recently highlighted by Poggio (2012), is to include development and learning in a full explanation of how sensory systems work. In the present case, we would like to know how development and learning shape humans' abilities to use multiple senses to detect nearby objects, and how these processes are reflected both in the changing computations that sensory systems carry out, and in changes to the neural circuits underlying these.

Acknowledgments

Thanks to Samir Zughni for help with programming. Supported by Economic and Social Research Council grants RES-062-23-0819, RES-061-25-0523 and ES/K005308/1, grant 220020240 from the James S. McDonnell Foundation 21st Century Science Scholar in Understanding Human Cognition Program, the NIHR Moorfields Biomedical Research Centre for Ophthalmology, and a Royal Society Wolfson Research Merit Award.

References

Alais, D. & Burr, D. (2004). The Ventriloquist Effect Results From Near Optimal Crossmodal Integration. *Current Biology*, 14, 257-262.

Bahrick, L. E. & Lickliter, R. (2000). Intersensory Redundancy Guides Attentional Selectivity and Perceptual Learning in Infancy. *Dev. Psychol.*, 36, 190-201.

Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The Race That Precedes Coactivation: Development of Multisensory Facilitation in Children. *Dev. Sci.*, 12, 464-473.

Barutchu, A., Crewther, S. G., Fifer, J., Shivdasani, M. N., Innes-Brown, H., Toohey, S., Danaher, J., & Paolini, A. G. (2011). The Relationship Between Multisensory Integration and IQ in Children. *Dev. Psychol.*, 47, 877-885.

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual Integration in Noise by Children and Adults. *J. Exp. Child Psychol.*, 105, 38-50.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The Physics of Optimal Decision Making: a Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks. *Psychol. Rev.*, 113, 700-765.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The Development of Audiovisual Multisensory Integration Across Childhood and Early Adolescence: a High-Density Electrical Mapping Study. *Cereb. Cortex*, 21, 1042-1055.

Carpenter, R. H. & Williams, M. L. (1995). Neural Computation of Log Likelihood in Control of Saccadic Eye Movements. *Nature*, 377, 59-62.

Clark, J. J. and Yuille, A. L. (1990). *Data Fusion for Sensory Information Processing Systems*. Boston, MA: Kluwer Academic

Diederich, A. (1995). Intersensory Facilitation of Reaction Time: Evaluation of Counter and Diffusion Coactivation Models. *Journal of Mathematical Psychology*, 39, 197-215.

Ernst, M. O. (2005) . A Bayesian view on multimodal cue integration. In *Human Body Perception from the Inside Out*, ed. Knoblich, G., Thornton, I. M., Grosjean, M., & Shiffrar, M. New York: Oxford University Press

Ernst, M. O. & Banks, M. S. (2002). Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion. *Nature*, 415, 429-433.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain Development During Childhood and Adolescence: a Longitudinal MRI Study. *Nat. Neurosci.*, 2, 861-863.

Gold, J. I. & Shadlen, M. N. (2002). Banburismus and the Brain: Decoding the Relationship Between Sensory Stimuli, Decisions, and Reward. *Neuron*, 36, 299-308.

Gold, J. I. & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annu. Rev. Neurosci.*, 30, 535-574.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, 18, 694-698.

Gori, M., Sandini, G., & Burr, D. (2012). Development of Visuo-Auditory Integration in Space and Time. *Front Integr. Neurosci.*, 6, 77.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination Rule for Redundant Information in Reaction Time Tasks With Divided Attention. *Percept. Psychophys.*, 35, 451-463.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-Auditory Interactions in Sensorimotor Processing: Saccades Versus Manual Responses. *J. Exp. Psychol. Hum. Percept. Perform.*, 20, 131-153.

Huttenlocher, P. R. & Dabholkar, A. S. (1997). Regional Differences in Synaptogenesis in Human Cerebral Cortex. *J Comp Neurol.*, 387, 167-178.

Lewkowicz, D. J. & Turkewitz, G. (1981). Intersensory Interaction in Newborns: Modification of Visual Preferences Following Exposure to Sound. *Child Dev.*, 52, 827-832.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory Interactions Within Human Primary Cortices Revealed by BOLD Dynamics. *Cereb. Cortex*, 17, 1672-1679.

Miller, J. (1982). Divided Attention: Evidence for Coactivation With Redundant Signals. *Cogn Psychol.*, 14, 247-279.

Miller, J. (1986). Timecourse of Coactivation in Bimodal Divided Attention. *Percept. Psychophys.*, 40, 331-343.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory Auditory-Visual Interactions During Early Sensory Processing in Humans: a High-Density Electrical Mapping Study. *Brain Res. Cogn Brain Res.*, 14, 115-128.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of Visual Cues Is Not Mandatory in Children. *Proc. Natl. Acad. Sci. U. S. A*, 107, 17041-17046.

Nardini, M., Begus, K., & Mareschal, D. (2013). Multisensory Uncertainty Reduction for Hand Localization in Children and Adults. *J. Exp. Psychol. Hum. Percept. Perform.*, 39, 773-787.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, 18, 689-693.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of Multisensory Spatial Integration and Perception in Humans. *Dev. Sci.*, 9, 454-464.

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual Decisions Formed by Accumulation of Audiovisual Evidence in Prefrontal Cortex. *J. Neurosci.*, 30, 7434-7446.

Otto, T. U. & Mamassian, P. (2012). Noise and Correlations in Parallel Perceptual Decision Making. *Curr. Biol.*, 22, 1391-1396.

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When Vision Is Not an Option: Children's Integration of Auditory and Haptic Information Is Suboptimal. *Dev. Sci.*, 17, 376-387.

Poggio, T. (2012). The Levels of Understanding Framework, Revised. *Perception*, 41, 1017-1023.

Raab, D. H. (1962). Statistical Facilitation of Simple Reaction Times. *Transactions of the New York Academy of Sciences*, 24, 574-590.

Reddi, B. A., Asrress, K. N., & Carpenter, R. H. (2003). Accuracy, Information, and Response Time in a Saccadic Decision Task. *J. Neurophysiol.*, 90, 3538-3546.

Schroger, E. & Widmann, A. (1998). Speeded Responses to Audiovisual Signal Changes Result From Bimodal Integration. *Psychophysiology*, 35, 755-759.

Schwarz, W. (1989). A New Model to Explain the Redundant-Signals Effect. *Percept. Psychophys.*, 46, 498-500.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge: MIT Press

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2009). The Neural Basis of Multisensory Integration in the Midbrain: Its Organization and Maturation. *Hear. Res.*, 258, 4-15.

Townsend, J. T. & Eidels, A. (2011). Workload Capacity Spaces : A Unified Methodology for Response Time Measures of Efficiency As Workload Is Varied. *Psychonomic Bulletin and Review*, 18, 659-681.

Townsend, J. T. & Nozawa, G. (1995). On the Spatio-Temporal Properties of Elementary Perception: An Investigation of Parallel, Serial, and Coactive Theories. *Journal of Mathematical Psychology*, 39, 321-359.

Townsend, J. T. & Wenger, M. J. (2004). A Theory of Interactive Parallel Processing: New Capacity Measures and Predictions for a Response Time Inequality Series. *Psychological Review*, 111, 1003-1035.

Wallace, M. T. & Stein, B. E. (1997). Development of Multisensory Neurons and Multisensory Integration in Cat Superior Colliculus. *J Neurosci.*, 17, 2429-2444.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and Integration of Multiple Sensory Inputs in Primate Superior Colliculus. *J. Neurophysiol.*, 76, 1246-1266.

Supplemental Methods

Supplemental Figures S1-2

Supplemental Tables S1-5

Supplemental Methods

Identification and exclusion of early responses

For each age group, we estimated the boundary of the "early" distribution by fitting bilinear regression lines to a normal plot of reciprocal response times (a recinormal plot). Since distributions of 1/reaction time tend to be normally distributed, on a recinormal plot the main distribution of responses tends to lie on a straight line, while early responses unrelated to the stimulus lie on another line with a shallower slope (Carpenter & Williams, 1995; Reddi, Asrress, & Carpenter, 2003). We estimated the boundary of the early distribution as the change-point of a bilinear fit to the earliest 25% of responses for each age group; see Supplemental Fig. S1. Based on this, the criteria for exclusion were <374ms at 4 years (3.4% of trials excluded), <335ms at 5-6 years (1.4 % excluded), <251ms at 7-8 years (0.9 % excluded), and <246ms at 9-10 years (0.9 % excluded). Adults did not show early responses (see Supplemental Fig. S1).

Classic race model

Given auditory only (A), visual only (V) and auditory-visual (AV) stimuli, cumulative distributions of response times in single-cue conditions A, V, state the probability p of a response time RT by time t, i.e. $p\,(RT \leq t|A)$ and $p\,(RT \leq t|V)$. Given parallel processing of the two signals A and V, and a strategy of responding to the fastest available signal on every trial, the probability of a decision before time t given simultaneous presentation of both signals (AV), is

$$p(RT \leq t|AV) = p(RT \leq t|A) + p(RT \leq t|V) - p(RT \leq t|A \cap RT \leq t|V) \quad \text{Eq. 1}$$

The final joint probability term depends on the correlation between the two channels' response times. Miller's model adopts the conservative approach of considering a correlation of -1, which predicts the greatest decreases in RT, and makes the joint probability (the third term in Eq. 1) zero.

We used the approach described in Ulrich, Miller, & Schroter (2007). For each participant, cumulative distributions of response times in single-cue conditions *A* and *V* were summed to give Miller's bound, *AVm*, for comparison with the empirical cumulative distribution of two-cue response times, *AV*. The *AV* and *AVm* distributions were compared at 10 deciles from 0.05 to 0.95. For each participant the cumulative reaction time associated with each decile was determined by linear interpolation. For each age group, at each decile, the hypothesis that empirical times (*AV*) are faster than times predicted by Miller's bound (*AVm*) was tested using a one-tailed paired t-test. To obtain a family-wise Type I error rate of 5% over all 10 such comparisons in each age group, the Šidák correction (Šidák, 1967) was applied, giving a significance criterion of 0.0051 per comparison:

$$1 - (1-\propto)^{1/n} \qquad \text{Eq. 2}$$

Where α is the familywise error rate (0.05) and n is the number of comparisons (10).

Extended race model

We followed the method developed by Otto & Mamassian, 2012; see the article for full details. For each participant, the cumulative distribution of response times in each of the conditions *A*, *V* and *AV* was fitted by a cumulative recinormal distribution by a least-squares fit with mean and SD as

free parameters. This is equivalent to fitting a cumulative normal distribution to the reciprocal of response times, i.e. 1/RT. The extended race model predicts cumulative distributions of two-cue response times for a participant always responding to the faster of two redundant sensory signals. This is based on independent processing of the two single channels, given two parameters, correlation between signals (ρ) and added noise (η) when processing both at once. Added noise states the proportion by which the standard deviation of each single cue's drift rate (in $1/t$ units) increases when both cues are processed together. Having fitted single-cue *A* and *V* cumulative response time distributions by cumulative recinormal distributions, we entered the means and SDs describing these into the extended race model with free parameters ρ, η in order to provide a least-squares fit to the empirical distribution of *AV* response times (examples, Fig. 4a, dashed lines).

An important feature of the approach of Otto & Mamassian (2012) is that model predictions are made in the $1/t$ domain. That is, the model predicts the effect of a race between signals on the *rate* (and variability in this rate) at which decisions are made (see Fig. 1e). We adopted the same approach to formulate a coactivation model.

Coactivation model

Our coactivation model (Fig. 1d) predicts cumulative distributions of two-cue response times for a participant responding based on pooled evidence from two redundant sensory signals. Coactivation models assume that multiple sources of evidence are pooled towards a single decision (Townsend & Wenger, 2004; Townsend & Eidels, 2011). In the present case, we have normally distributed distributions of response rates (in $1/t$ units) when subjects are given either single cue A or V. We would like to predict their distribution of response rates when both are available together (AV) if the A and V information is being pooled. The simplest model would be one in which A and V rates are uncorrelated, and the pooling mechanism is perfectly efficient. In that case, the distribution of AV rates would be the sum of two independent normally distributed variables, where the mean is the sum of the two means, and the variance is the sum of the two variances:

$$\mu_{AV} = \mu_A + \mu_V \quad \text{Eq. 3}$$

$$\sigma_{AV}^2 = \sigma_A^2 + \sigma_V^2 \quad \text{Eq. 4}$$

In practice, increases in the AV rate were lower, and increases in the variance of this rate were also lower, than this parameter-free model would predict.

We allowed that A and V rates might be correlated, and that the pooling of the two signals may not be perfectly efficient. Our prediction for the variance in AV rate, allowing for a possible correlation between A and V rates, is the total variance for a sum of two correlated normally distributed variables:

$$\sigma_{AV}^2 = \sigma_A^2 + \sigma_V^2 + 2\rho \cdot \sigma_A \sigma_V \quad \text{Eq. 5}$$

Where ρ is the correlation between A and V rates, and must range from -1 to 1. The total variance is lower when distributions are negatively correlated than when there is no correlation, or a positive correlation.

Our prediction for the mean AV rate, allowing that the accumulation of combined evidence from A and V stimuli may not be perfectly efficient, includes an efficiency parameter η. This states the proportion of the theoretically optimal total $A + V$ rate that is in practice achieved when A and V signals are pooled:

$$\mu_{AV} = (\mu_A + \mu_V) \cdot \eta \qquad \text{Eq. 6}$$

Where η must range from 0 to 1.

References

Carpenter, R. H. & Williams, M. L. (1995). Neural Computation of Log Likelihood in Control of Saccadic Eye Movements. *Nature*, 377, 59-62.

Otto, T. U. & Mamassian, P. (2012). Noise and Correlations in Parallel Perceptual Decision Making. *Curr. Biol.*, 22, 1391-1396.

Reddi, B. A., Asrress, K. N., & Carpenter, R. H. (2003). Accuracy, Information, and Response Time in a Saccadic Decision Task. *J. Neurophysiol.*, 90, 3538-3546.

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *J Am Stat Assoc*, 62, 626-633.

Townsend, J. T. & Eidels, A. (2011). Workload Capacity Spaces : A Unified Methodology for Response Time Measures of Efficiency As Workload Is Varied. *Psychonomic Bulletin and Review*, 18, 659-681.

Townsend, J. T. & Wenger, M. J. (2004). A Theory of Interactive Parallel Processing: New Capacity Measures and Predictions for a Response Time Inequality Series. *Psychological Review*, 111, 1003-1035.

Ulrich, R., Miller, J., & Schroter, H. (2007). Testing the Race Model Inequality: an Algorithm and Computer Programs. *Behav. Res. Methods*, 39, 291-302.

Supplemental Figures

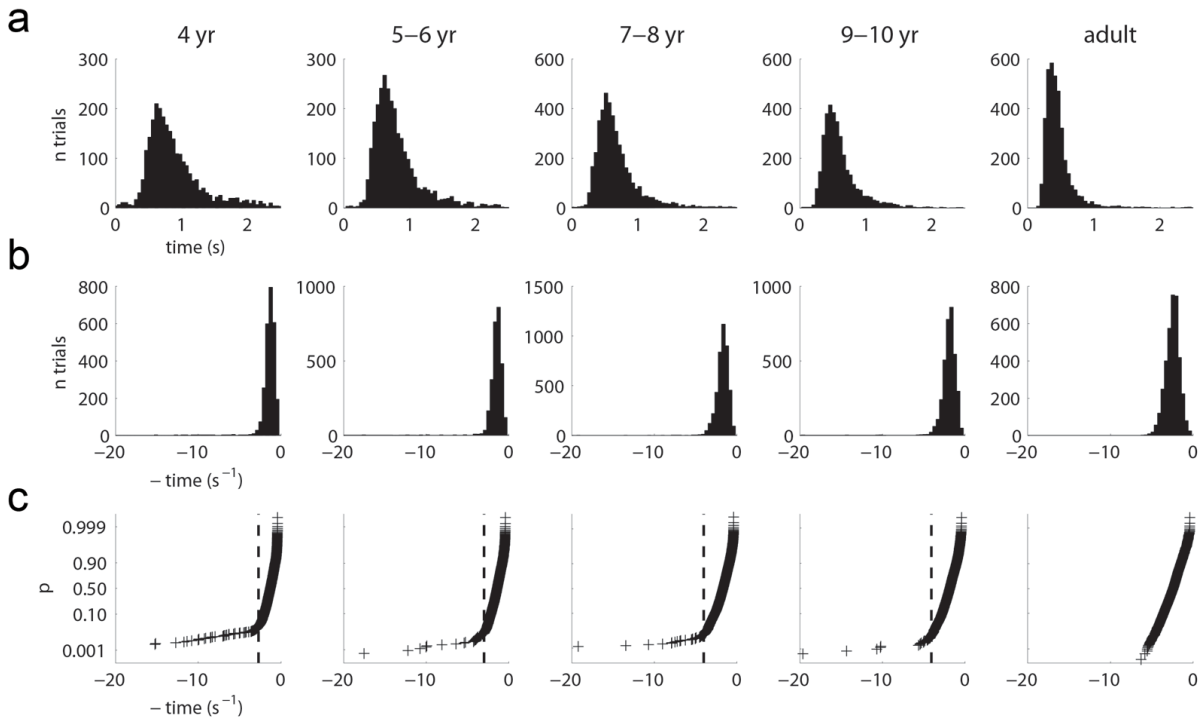

Figure S1. Response time distributions and cut-offs for "early" responses. **a.** distributions of response times, pooled across all three experimental conditions. **b.** distributions of negative reciprocal response times. **c.** normal plots of negative reciprocal response times. Dashed lines indicate change-points of bilinear fits, used as a cut-off to define early responses.

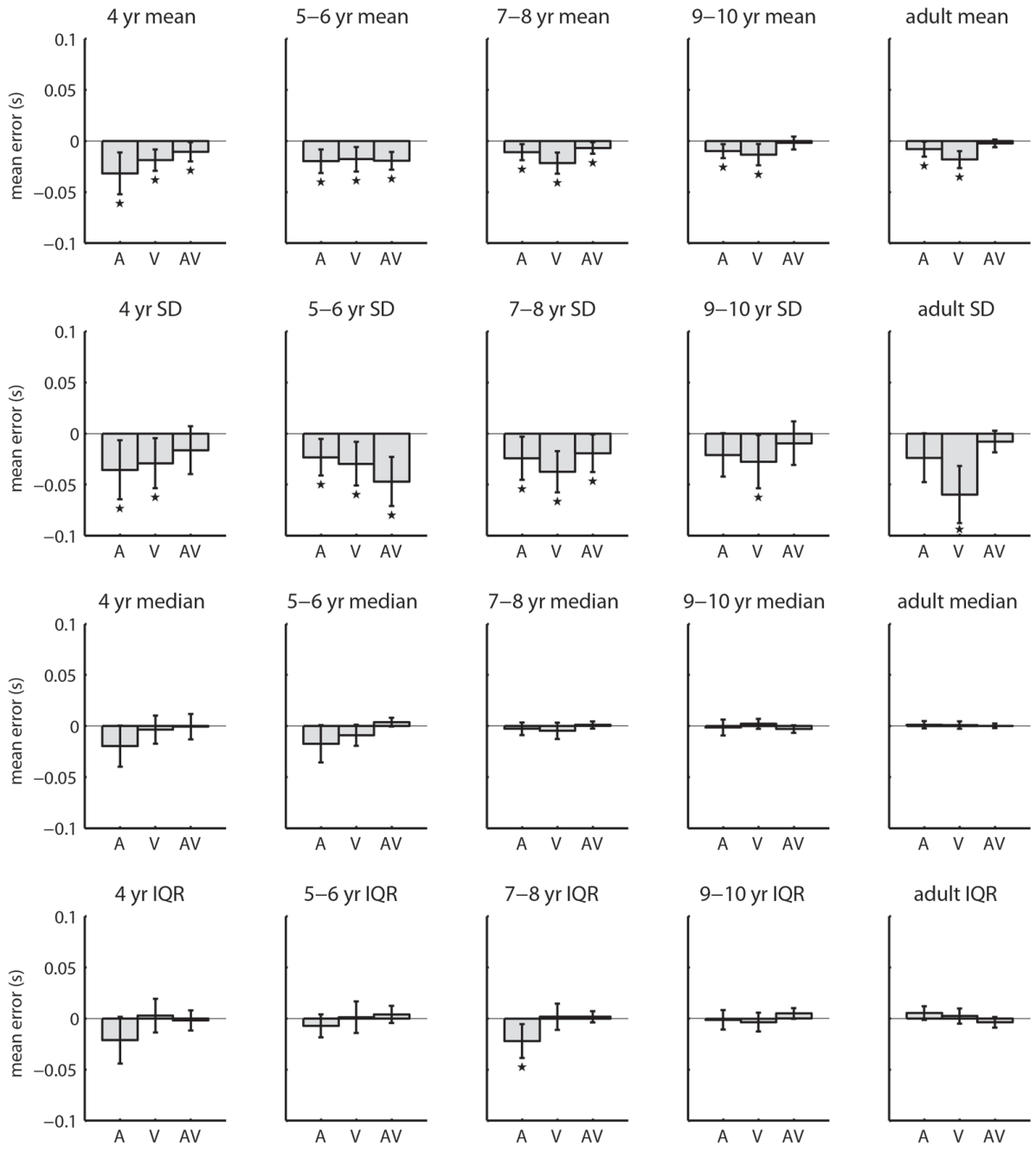

Figure S2. Mean ± 95% c.i. error in estimates of mean, SD, median, and IQR of response times via recinormal fits to reaction time data in the extended race model. * 95% confidence interval excludes zero.

Supplemental Tables

Table S1. Results of planned comparisons (two-tailed paired t-tests) for mean response times in condition *AV vs.* conditions *A* and *V*; see Fig. 1b-c. * significant at the 5% level.

measure and age group	***AV vs A***	***AV vs V***
mean, 4 yr	t(16) = 5.8, p < 0.001 *	t(16) = 8.5, p < 0.001 *
mean, 5-6 yr	t(20) = 7.0, p < 0.001 *	t(20) = 8.1, p < 0.001 *
mean, 7-8 yr	t(31) = 15.5, p < 0.001 *	t(31) = 13.5, p < 0.001 *
mean, 9-11 yr	t(24) = 10.2, p < 0.001 *	t(24) = 8.4, p < 0.001 *
mean, adult	t(25) = 14.9 p < 0.001 *	t(25) = 13.1, p < 0.001 *
SD, 4 yr	t(16) = 3.6, p < 0.01 *	t(16) = 4.6, p < 0.001 *
SD, 5-6 yr	t(20) = 2.6, p < 0.02 *	t(20) = 4.4, p < 0.001 *
SD, 7-8 yr	t(31) = 4.9, p < 0.001 *	t(31) = 6.5, p < 0.001 *
SD, 9-11 yr	t(24) = 4.4, p < 0.001 *	t(24) = 4.6, p < 0.001 *
SD, adult	t(25) = 4.5, p < 0.001 *	t(25) = 6.7, p < 0.001 *

Table S2. Results of one-tailed paired t-tests comparing cumulative response time deciles with race model predictions (Miller's bound); see Fig. 1d. * significant at the 5% level after Šidák correction for multiple comparisons.

		age group				
		4 yr	***5-6 yr***	***7-8 yr***	***9-10 yr***	***adult***
decile	***d.f.***	16	20	31	24	25
0.05	***t, p***	3.5, <0.002 *	4.9, <0.001 *	5.1, <0.001 *	5.4, <0.001 *	14.2, <0.001 *
0.15	***t, p***	3.5, <0.002 *	5.7, <0.001 *	7.7, <0.001 *	7.5, <0.001 *	19.3, <0.001 *
0.25	***t, p***	2.4, <0.02	6.0, <0.001 *	8.6, <0.001 *	5.9, <0.001 *	18.4, <0.001 *
0.35	***t, p***	2.5, <0.02	4.5, <0.001 *	10.1, <0.001 *	6.0, <0.001 *	14.2, <0.001 *
0.45	***t, p***	1.0, 0.16	4.0, <0.001 *	8.0, <0.001 *	6.5, <0.001 *	13.0, <0.001 *
0.55	***t, p***	-0.6	2.5, <0.02	4.7, <0.001 *	4.6, <0.001 *	12.4, <0.001 *
0.65	***t, p***	-3.1	0.6, 0.27	0.6, 0.28	2.1, <0.05	10.8, <0.001 *
0.75	***t, p***	-3.9	-1.3	-1.9	0.7, 0.24	4.0, <0.001 *
0.85	***t, p***	-6.3	-3.9	-4.8	-2.1	-0.5
0.95	***t, p***	-7.3	-6.6	-8.3	-5.8	-6.1

Table S3. Mean (s.e.) parameters of fits of reci-normal distributions to empirical data.

group	μ_A (s^{-1})	μ_V (s^{-1})	μ_{AV} (s^{-1})	σ_A (s^{-1})	σ_V (s^{-1})	σ_{AV} (s^{-1})
4 year	1.255 (0.075)	1.195 (0.059)	1.476 (0.069)	0.367 (0.024)	0.355 (0.018)	0.376 (0.021)
5-6 year	1.295 (0.064)	1.318 (0.047)	1.673 (0.060)	0.361 (0.019)	0.424 (0.020)	0.416 (0.025)
7-8 year	1.598 (0.059)	1.532 (0.060)	2.022 (0.077)	0.427 (0.022)	0.491 (0.020)	0.503 (0.024)
9-11 year	1.639 (0.071)	1.724 (0.056)	2.230 (0.083)	0.410 (0.018)	0.487 (0.019)	0.521 (0.032)
adult	2.214 (0.077)	2.242 (0.074)	3.024 (0.105)	0.493 (0.026)	0.532 (0.021)	0.628 (0.029)

Table S4. Mean (SE) of differences in empirical and model-predicted median and IQR of response times for condition *AV*, with extended race model fits minimising square error of the fit to cumulative *AV* reaction times, and results of two-tailed one-sample t-tests (see Fig. 3).
* significant at the 5% level.

age group	***median, data vs model***		***IQR, data vs model***	
	mean (SE) (ms)	***t-test***	***mean (SE) (ms)***	***t-test***
4 yr	-3.46 (1.17)	$t(16) = 3.0$, $p < 0.01$ *	-10.62 (5.39)	$t(16) = 2.0$, $p = 0.07$
5-6 yr	-4.15 (1.68)	$t(20) = 2.5$, $p < 0.05$ *	-8.85 (4.19)	$t(20) = 2.1$, $p < 0.05$ *
7-8 yr	-3.73 (0.59)	$t(31) = 6.4$, $p < 0.001$ *	-5.84 (2.12)	$t(31) = 2.8$, $p < 0.01$ *
9-11 yr	-2.47 (3.02)	$t(24) = 0.8$, $p = 0.42$	-14.09 (4.64)	$t(24) = 3.0$, $p < 0.01$ *
adult	-5.09 (0.50)	$t(25) = 10.2$, $p < 0.001$ *	-15.15 (2.42)	$t(25) = 6.3$, $p < 0.001$ *

Table S5. Results of paired two-tailed t-tests comparing MSEs of fits to empirical data by the extended race model and the coactivation model.

age group	***MSE, race vs coactivation model***
4 yr	$t(15) = 2.6$, $p < 0.03$ *
5-6 yr	$t(20) = 2.6$, $p < 0.02$ *
7-8 yr	$t(31) = 2.8$, $p < 0.01$ *
9-11 yr	$t(24) = 2.9$, $p < 0.01$ *
Adult	$t(25) = 4.7$, $p < 0.001$ *